DiffuserCam: lensless single-exposure 3D imaging using Custom Hardware

Commented [LB1]: 1. Replace figures from Antipa paper with your own (Use PPT or Illustrator to create a diagram)
2. Switch fig 1 & 2. Top level flow goes first.
3. No need black box for each figure
4. Change the figures to half column wide.
3. Incorporate all the figures (Figure 1-11) in the main manuscript with max 6 pages.
4. Equation should be typed in.

Ray Duran
Department of Electrical Engineering and Computer Science
Univertisy of North Dakota
ray.duran@und.edu

Abstract— **In the paper by Antipa et al. [1], a lens less diffuser cam for 3-D imaging is introduced. Using a unique pseudorandom pattern of caustics on the sensor, simple calibration, and computational processing we solve for 3-D voxels with a single exposure. The use of compressive sensing and a numerical method allows us to solve the matrix inversion problem without dealing with the large original matrix. The simplified convolution forward model, validated experimentally, has a FoV of +/- 42 degrees in the x -axis and +/- 30.5 degrees in y-axis and is object dependent for resolution as reported in [2] (Supplementary Antipa). We are constructing a custom enclosure for our CMOS sensor that includes a diffuser and aperture to duplicate the results in [1]. We also followed the instructions in a tutorial reference by [1], where we used commodity parts and homemade optics where we were not able to duplicate experimental results.**

I. Background

The main motivation for a DiffuserCam is to provide a light field representation of 4-D information into a 2-D sensor. While capturing spatial and angle information we want the system to be inexpensive, flexible, and compact. Providing such a system as described above falls in the realm of light-field imaging which lends itself to many applications: 3D neural activity[3],compressive radar imaging[4] , synthetic aperture[5], and visual odometry[6].

Why diffusers for lensless systems? At first glance a diffuser is a highly diffractive medium that would not seem to lend itself to light field imaging. A coded aperture light field is an alternative to a diffuser that can also achieve higher resolution than conventional cameras [7]. However, an amplitude mask system by its' very nature limits the amount of light that the sensor can record. A micro lens array can gather more information that just x-y coordinates but suffers from scaling.

One of the key theoretical pillars in being able to use diffusers in a lensless system is that the surface of the diffuser can be modeled as a smooth Gaussian smooth surface. So the diffractive effect causing speckles (caustic patterns) from interference can be used for reconstruction. This is because each area of the diffuser is unique creating a type of signature that can encode plenoptic information about our illuminated object. These properties, of uniqueness, in theory allow us a better scaling that can be utilized by a micro lens array system. Another important point is that diffusers which are essentially phase shifter elements can concentrate the light from a 4-D light field into a 2-D sensor better than an amplitude mask using Fourier optics.

For the reasons above, a diffuser allows an encoding of the light field beyond just x-y coordinates and because of the random surface of a diffuser, modeling the recorded light allows

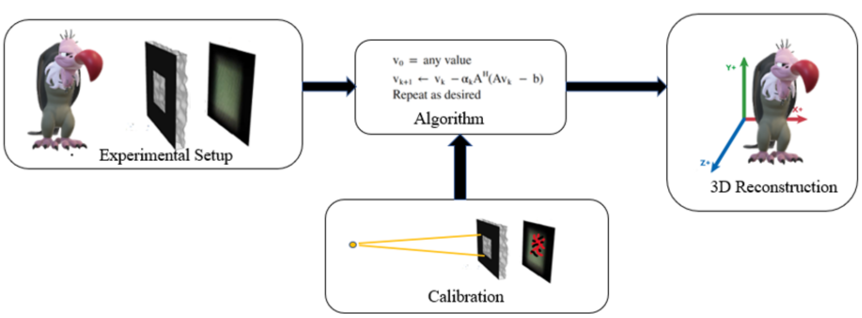

Fig. 1. Top level flow of diffuser cam

us linearity in reconstruction that lends itself to well-established inverse problem definition and optimization techniques.

An important point to using diffusers and the randomness they provide is to see if this fits within the theoretical framework of compressive sensing, that has conditions on the inverse matrix. When trying to solve diffuser inverse problems with less than full rank, it is possible to recover the original vector if some conditions on a matrix are met. Such properties if satisfied such as Full Spark of matrix, and Null Space Property, both have full rank submatrices. A less computationally intensive way to calculate the sparsity that can be recovered is by computing the mutual coherence of the matrix, the combination of the measurement and representation matrix. Here is this paper we forego the formality of the mathematics and assume that our 3D objects are sparse in some domain.

A final point to consider when constructing our diffuser lensless system is to consider how a 3-D object will affect the caustic patterns on our sensor. Fig. 1 shows the effect of lateral, and depth shifts onto the sensor. Fig 2. is a top-level flow of system.

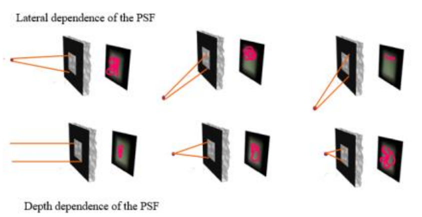

Fig. 2. Lateral and depth dependencies of PSF thru diffuser

II. Methods

In theory a lens less system has all the information available at the sensor. However, the problem is ill-posed. Introducing a diffuser gives structure to our point spread function, or random matrix, allowing us to solve an otherwise intractable problem. If we enforce sparsity as a prior, and non-negativity (no negative pixels), minimize for the least squares we are trying to solve the equation below

We need to solve the minimization problem of:

$$g(v) = \frac{1}{2}\|Av - b\|_2^2 \qquad (1)$$

$$\nabla_v g(v) = A^H(Av - b) \qquad (2)$$

Here in (1) we do not show the regularization term that helps enforce sparsity, instead we want to show the main thrust of how we approach minimization overall

$$Av = CMv \qquad (3)$$

$$A^H v = M^H C^H v \qquad (4)$$

Here, the significance of (4) is that we have reduced the problem of computing A which could be a very large matrix to the adjunct of M and C, a much more manageable problem. (C is introduced as a cropping matrix)

$$M^H v = (F^{-1}\ \mathrm{diag}(Fh)\ F)^H v \qquad (5)$$

$$= (F^H\ \mathrm{diag}(Fh)^H\ (F^{-1})^H) v \qquad (6)$$

$$= F^H\ \mathrm{diag}(Fh)^*\ F(v) \qquad (7)$$

Here there are a few points to note. The first is that the M matrix in (5) has already been decoupled from A, and this is because if we included the cropping effect into A, the matrix would be ill-conditioned. Continuing with (5), we have just expressed the convolution of "Mv" as a product of Fourier and the inverse of that dot product but have expressed the equation as products of matrices.

$$A = CF^{-1}\ \mathrm{diag}(Fh)\ F \qquad (8)$$

$$A^H = F^{-1}\ \mathrm{diag}(Fh)^*\ FC^H \qquad (9)$$

Here we substituted (7) back into (4) to give us (8) and taken the adjunct of (8) to give us (9). At this point (8) and (9) we can implement into code to solve our iteration (10).

Formatted: English (United States)

$$v_0 = \text{any value}$$

$$v_{k+1} \leftarrow v_k - \alpha_k A^H(Av_k - b) \qquad (10)$$

Repeat as desired

For our experiments we used a Raspberry Pi Camera V2 sensor with some modifications (see discussion section). The sensor itself is a CMOS device with 3280(H) x 2464(V) active pixel count. The platform system is a Raspberry Pi 4 Model B with HDMI ports and an interface for an external sensor. For our light source we used a flashlight with a pinhole covering the LED and black adhesive to limit output other than a pin hole to simulate our point source for calibration. Another flashlight was used to properly illuminate our object of interest.

The diffusers we used were single and double-sided ordinary scotch tape. We also constructed with two paper clips that supported a mask that fit over the CMOS sensor assembly, see Fig. 3. The entire assembly was outfitted with black opaque tape to block and limit stray light other than from our point source. All the experiments were done in as dark an environment as possible. Black tape was also used for an aperture.

Fig. 3. Aperture, sensor, and makeshift housing to hold diffuser over sensor

We did not perform a proper calibration (see discussion for more explanation). We estimated the focal distance for our camera to just a few millimeters.

III. Results

All results are using the FISTA algorithm, a speedup of the gradient method that was outlined in the previous section. The average computation time for reconstruction is about 30 seconds for 100 iterations. This is run on a Dell workstation that uses an Intel Core i7-8700 CPU @ 3.2Ghz with 32.0GB of RAM using Windows 10 Enterprise operating system. All code is run on PyCharm 2020.1.1 Professional Edition. Python code is downloaded and slightly modified to run on our workstation environment from [1].

The object that was used for all experiments is a pen with the face of an animated character, see Fig 10. Fig 4. Shows our results from using double-sided tape. The algorithm for all experiments was run with at least 80 iterations using the Fast Iterative Shrinkage Thresholding Algorithm (FISTA) instead of a simple gradient. Also, for even faster speed there are options to use Alternating Direction Method of Multipliers (ADMM). Results for single sided, Antipa and double-sided response are in Figs. 4-9 respectively.

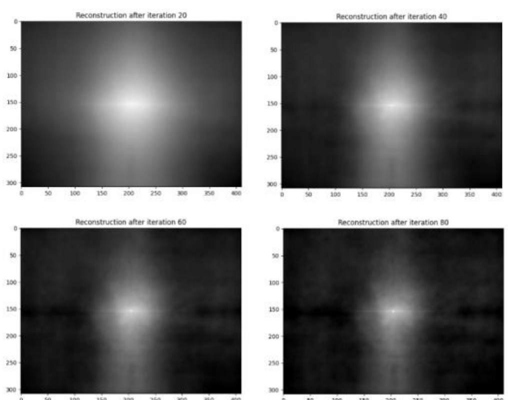

Fig. 4. Double sided tape used as a diffuser. Upper Left is for 20 iterations reconstructions. Upper right, Lower left, and Lower right are 40,60 and 80 iterations of FISTA algorithm

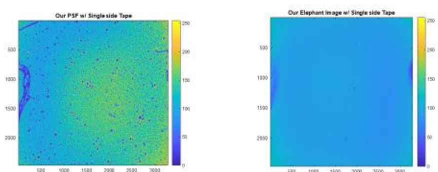

Fig. 5. Single sided tape PSF (left) and captured response of object(right).

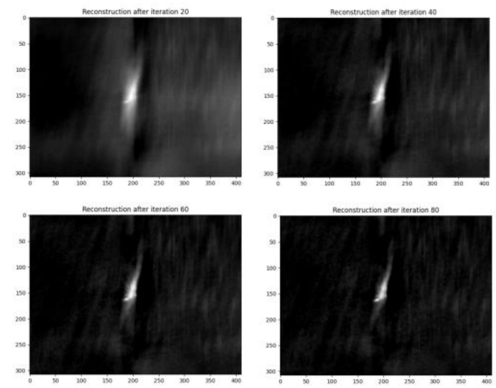

Fig. 6. Single sided tape used as a diffuser. Upper Left is for 20 iterations reconstructions. Upper right, Lower left, and Lower right are 40,60 and 80 iterations of FISTA algorithm

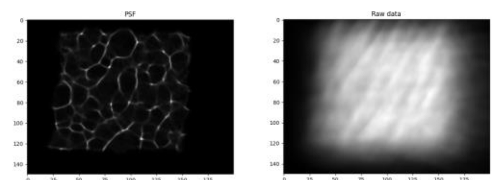

Fig. 7. From Antipa [1] PSF (left) and captured response of object(right).

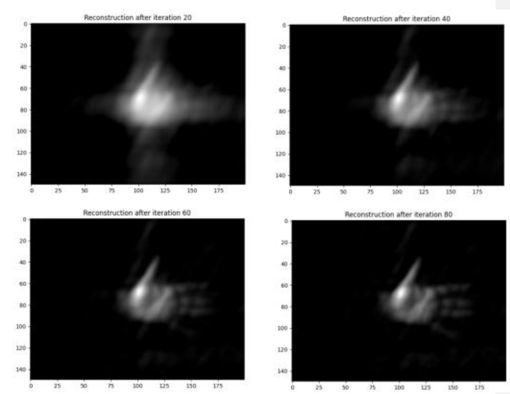

Fig. 8. Antipa [1] diffuser. Upper Left is for 20 iterations reconstructions. Upper right, Lower left, and Lower right are 40,60 and 80 iterations of FISTA algorithm

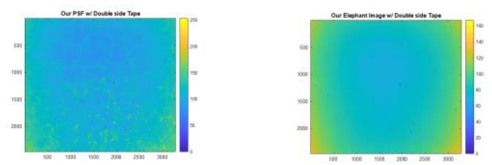

Fig. 9. Double sided tape PSF (left) and captured response of object(right).

Fig. 10. Object used for single and doubled sided tape experiments (pink elephant pen)

IV. DISCUSSION

In the original paper [1], there is a variation where the equipment used is commodity. The construction of a system while inexpensive and convenient is not practical and stable to construct consistently with success. The author followed the instructions outlined in a tutorial and had difficulty recreating the caustic pattern point spread function.

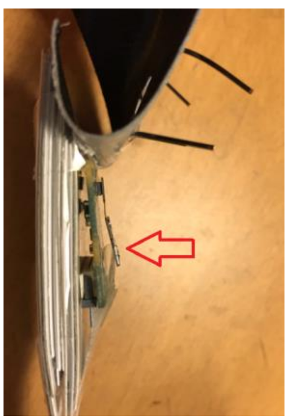

Fig. 11. Red arrow shows tilt to sensor assembly

One example of a stability and repeatable solution is the conversion of a raspberry sensor for measurements. The first problem with using a sensor like the raspberry pi sensor is that the sensor is not a flat sensor and requires that the attached lens be removed to place the diffuser a few millimeters away to produce a good caustic pattern(see Fig. 11.). However, this requires some disassembly that is not recommended by manufacturer and if not done carefully could damage the sensor. A second problem is how to determine calibration. It is challenging to hold the diffuser so close, a few millimeters away from the CMOS sensor of the raspberry pi. Another difficult problem is that even after we have estimated the location that we want our diffuser to be from the sensor, it was still demanding to place the whole assembly at the right distances for the sensor, diffuser, and aperture (see Fig. 12.) A final challenge is how to with a non-varying light source, calibrate the PSF at a distance from then sensor for a good reading.

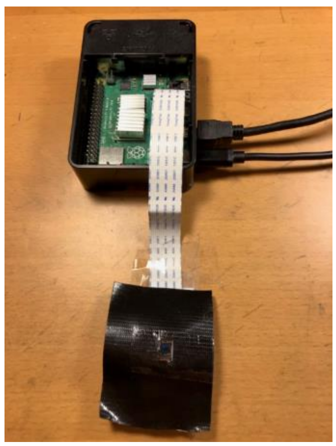

Fig. 12. Enclosed: sensor, diffuser and aperture

As a result, with a small investment, several hundred dollars, we propose a more stable and practical solution that increases the prospect of repeated success. As a first step we construct a custom housing using a 3D printer. The housing is built on top of a sensor. The housing could stabilize the sensor and with a way to move the aperture up and down provide an easy way to perform the calibration that would yield us a good point spread function that would be manifested as an expected caustic pattern.

There are some other latent issues with the diffuser cam: Field-of-view (FoV), off-axis modeling, number of calibrations needed, resolution and finally the complexity of the object.

The first, FoV, has an angular response with incident light where the sensor will not respond. This means that when we calculate the overall Fov equation we need to account for a cutoff angle to address angles with high incidence on our diffuser. Here with “good optics”, meaning not with commodity items that we used in our experiments a FoV of +/- 42 degrees in the x -axis and +/- 30.5 degrees in y-axis was attained. For off-axis modeling, when we develop a forward model, we often assume that the diffuser is a pure phase modulation element. The light field that we approximate then must be solved for phase distribution only. This of course is an assumption that means that there are no effects from off-axis. A third concern with using a diffuser cam lensless system is the number of calibrations needed for good reconstruction. Here, Antipa [1] states that only one calibration is needed, yet they even recommend one at each depth of the object, which adds complexity to the overall system and makes reconstruction problematic for objects that are complex. Finally, the number of point sources of the object limit the accuracy of reconstruction. Note, this is different than a regular camera where the computation is not a function of the object. Antipa [1] and Cai [8] both show the effect of increasing the complexity of and object and increase of the condition number for an increase in angular sampling.

V. Conclusion

Following the guide outlined in Antipa [1], we tried to build with the same components that were outlined in their tutorial. We listed two of the challenges encountered, one was modifying the sensor and the second was performing calibration. We were not able to get a satisfactory caustic pattern that would allow us with some accuracy to recover the object. As can be clearly seen from Fig. 5. and Fig. 6. (Left side of both figures) the PSF caustic patterns did not have the familiar high frequency ring that we were expecting. It is this high frequency PSF that allows us to capture many of the details in the target that we are trying to reconstruct. What we see in Fig 5 (right side) is a kind of low frequency response of the image. As an example of a working system, Fig. 7. shows the PSF and captured pattern respectively. Also, with both of our chosen diffusers, single sided tape, and double-sided tape, there were some other materials that were considered transparent, that we experimented and not listed here with worse results. These other materials had a similar point spread function recorded on our sensor.

There are four things that we would like to expand on after this project. The first is to develop a way to rapidly perform PSF measurements from various depths and off-axis with a minimal of preparation and with automation if possible. Here we would like to include the use of a digital micrometer device (DMD) to be included in the light path. The system would then comprise a laser directed at a DMD that would then be aimed at a small pin hole aperture. From the aperture, the light ray would strike our diffuser at different angles.

The second item that we hope to explore is to model PSFs with a formula like Jin [7] and look at off-axis performance. Jin outlines a method that can help with avoiding the use of extensive calibration to provide better reconstruction. The idea is based on using Fourier optics first to express the system PSF as a function with only one unknown variable, which is the diffuser phase distribution. Then Jin proceeds to estimate the diffuser phase by using laser beam shaping theory. Using an architecture with a Fourier lens between the diffuser and the sensor, and that fact that a far-field distribution of laser beams is proportional to the Fourier Transform of the near-field distribution, the phase of the diffuser corresponds to the phase of the diffuser which then corresponds to the near-field component, retrievable by phase retrieval methods [9].

The third area that we would like to expand on is how the models are built around light field fundamentals. For example, our diffuser cam using the ideas outlined by Antipa [1] is built by using ray optics. Here we note the main limitation of this approach is that it assumes that the diffuser to sensor distances is small. For many applications this is not a good assumption. An alternative way would use wave-optics that better model diffractive effects. Here the Wigner distribution, originally developed to study quantum mechanics, can be used to model the light field without the diffuser to sensor constraint [10].

The fourth and final area that we would like to consider after this project is how to speedup our processing so that the system with a diffuser can be used in real time. Here as a possible example, we can take algorithms used for reconstruction and try to implement this in a Field Programmable Gate Array (FPGA) [11]. Using an FPGA, to gain speed in processing, we can also try to gain speed by implementing an approach that scales well with the size of the linear problem. Solving the compressive sensing problem with Alternating Direction Method of Multipliers [12] instead of with other convex optimization methods, like enforcing an L1 norm as a regularizing term, allow us to use pre-computed portions of a LU decomposition [13]. In addition to using embedded or external memory of an FPGA for acceleration of our linear system, we can also use the built-in digital signal processing units for additional speed up as well as to implement a Floating Point Unit (FPU) in an FPGA to increased the overall accuracy of our results.

The overall paradigm of using a diffuser to capture 4-D plenoptic information in a 2-D sensor in engineering terms means less samples and an overall reduction in the complexity of the acquisition system. This translates into 1) less power, 2) less hardware ,3) simpler algorithms for acquiring (directly proportional to hardware), 4) less speed for sampling (even non-uniform sampling), 5) less bandwidth, and 6) less storage and a general ability of the engineer to trade-off performance given the “compactness” of sensing.

Further developing diffuser light field modeling to the point where complex objects can be reconstructed with robust and repeatable results will allow the realization of many new and interesting computational imaging applications.

References

[1] N. Antipa et al., “DiffuserCam: lensless single-exposure 3D imaging single-shot,” Optica 5,1-9 (2018).

[2] [N. Antipa et al., “DiffuserCam: lensless single-exposure 3D imaging single-shot,” Supplementary (2018).

[3] N.C. Pegard et al., “Compressive light-field microscopy for 3D neural activity recording,” Optica 3, 517-524 (2016).

[4] R. Baraniuk, P.Steeghs, “Compressive radar imaging,” IEEE Radar Conference (2007).

[5] V. Vaish, et al., “Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures,” Proc. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (IEEE, New York, 2006).

[6] N. Zeller, F. Quint, and U. Stilla, "From the calibration of a light-field camera to direct plenoptic odometry," IEEE J. Sel Top. Signal Process 11, 1004-1019 (2017).

[7] X. Jin, D. Mao,S. Wei,Q. Dai "Point spread function for diffuser camera based on wave propagation and projection model," Opt. Express 27, 12748-12761 (2019).

[8] M. Cai, J. Chen, G. Pedrini, W. Osten, X. Liu, and X. Peng, "Lensless light-field imaging through diffuser encoding,"Light Science & Applications (2020).

[9] J.R. Fienup, "Phase retreival algorithms: a comparison," Appl. Opt 21(15), 2758-2769 (1982).

[10] Z. Zhang, M. Levoy, " Wigner Distributions and how they relate to the light field,"Proceedings of IEEE ICCP (2009).

[11] J. L. Jerez, " PhD Thesis : Custom optimization algorithms for efficient hardware implenentation", Imperial College London, Electrical Engineering Dept., (2013)

[12] R. Hayakawa, "Asymptotic Analysis of ADMM for Compressed Sensing," Cornell University (2020)

[13] A. Abmann, Y. Wu, B. Stewart, A. Wallace, "Accelerated 3D image reconstruction for resource constrained systems," EUSIPCO (2020)